ULTRASOUND SPINE IMAGE SEGMENTATION USING MULTI-SCALE FEATURE FUSION SKIP-INCEPTION U-NET (SIU-NET)

Abstract— Scoliosis is a 3D spinal deformation where the spine takes a lateral curvature, forming an angle in the coronal plane. Diagnosis of scoliosis requires periodic detection, and frequent exposure to radiative imaging may cause cancer. A safer and more economical alternative imaging, i.e., 3D ultrasound imaging modality, is being explored. However, unlike other radiative modalities, an ultrasound image is noisy, which often suppresses the image's useful information. Through this research, a novel hybridized CNN architecture, multi-scale feature fusion Skip-Inception U-Net (SIU-Net), is proposed for a fully automatic bony feature detection, which can be further used to assess the severity of scoliosis safely and automatically. The proposed architecture, SIU-Net, incorporates two novel features into the basic U-Net architecture: (a) an improvised Inception block and (b) newly designed decoder-side dense skip pathways. The proposed model is tested on 109 spine ultrasound image datasets. The architecture is evaluated using the popular (i) Jaccard Index (ii) Dice Coefficient and (iii) Euclidean distance, and compared with (a) the basic U-net segmentation model, (b) a more evolved UNet++ model, and (c) a newly developed MultiResUNet model. The results show that SIU-Net gives the clearest segmentation output, especially in the important regions of interest such as thoracic and lumbar bony features. The method also gives the highest average Jaccard score of 0.781 and Dice score of 0.883 and the lowest histogram Euclidean distance of 0.011 than the other three models. SIU-Net looks promising to meet the objectives of a fully automatic scoliosis detection system.

Index Terms— Bony Feature, Convolutional Neural Network, segmentation, Scoliosis, Ultrasound, U-Net, Feature Fusion.

I.INTRODUCTION

Scoliosis is an ailment in which the spinal cord deforms progressively with time. Teenagers are at the highest risk since their skeletal structure is not fully developed. In medical parlance, this condition is known as Adolescent Idiopathic Scoliosis (AIS) [1]. If unattended, AIS can progressively deteriorate into severe physiological problems such as back pain, compression onto nerves, heart and lungs which can be a general impediment to the good health of the adolescent population [2],[3], [4]. The technique to detect and diagnose scoliosis is well established. Usually, the detection process entails scanning the spinal region of a patient using an appropriate modality and assessing the spine bend. The process is repeated several times over the patient's lifespan and if, at any stage, the bend of spine measures >10°, the patient is earmarked for potential scoliosis treatment.

The popular method of detection and diagnosis involves (a) multiple X-ray scanning sessions and (b) measurement of Cobb angle, which is a gold standard to assess AIS [5]. The main disadvantage of this method is that repeated exposure to ionizing radiation increases the risk of cancer [6]. Further, other non-ionizing radiation modalities, such as MRI, are very expensive and not readily designed for scanning in standing posture. Instead, to distinguish and study the bony features in the spine, imaging techniques, such as ultrasound, are being explored. Being non-radiating, ultrasound is not only safe but also affordable and quick [7].

Table 1 summarizes the various applications of ultrasound imaging in the diagnosis of spinal deformation. According to Hwang et al., the application of ultrasound imaging modality in the assessment of spinal health has been a popular subject of research in recent times [8]. Tawfik et al. [9] established that, in identifying main spinal abnormalities in infants, the diagnostic value of spinal ultrasonography was equivalent to that of MRI. In another research, Zhang et al. [10] showed that, for long-term spinal deformity treatment, such as cases of severe scoliosis, ultrasound imaging is safe and gives patient comfort along with a clear intrathecal structure for guided treatment. Ultrasound is also applied in pre-procedural imaging of central neuraxial blockade of the spine to identify, in real-time, the ideal trajectory (best angle, direction of approach, and depth) and to optimize the subsequent invasive treatment with fewer needle passes and skin punctures [11].

Despite the advantages of ultrasound imaging modality, an ultrasound image is difficult to handle. In conventional modalities such as X-ray and MRI, high-quality images are obtained and the bony features are relatively clear. However, an ultrasound image suffers from additional challenges such as (i) speckle noise (Fig. 1 (b) & (c)) and (ii) low contrast.

TABLE 1
APPLICATION OF ULTRASOUND IMAGING IN DIAGNOSIS OF SPINAL DEFORMITIES

Author	Field of work	Region of interest (ROI)	Comments
Hwang et al. 2021 [8]	Application of ultrasound in traumatic spinal cord injury (SCI)	Spinal cord	Ultrasound imaging can serve as a powerful adjunct to various developing therapies for SCI
Tawfik et al. 2020 [9]	Comparison of spinal ultrasound with MRI for the diagnosis of spinal deviation in infants	Spinal cord and bony elements	Spinal ultrasound can be used as a first-line screening investigation for infants with spinal deviation
Zhang et al. 2021 [10]	Nusinersen through lumbar puncture with real-time ultrasound guidance in spinal muscular atrophy (SMA) patients with severe scoliosis.	Lumbar area	100% success rate (no major complications) has been achieved in radiation-free and real-time ultrasound-guided lumbar intrathecal administration of nusinersen in SMA patients with severe scoliosis
Kalagara et al. 2021 [11]	Central neuraxial blockade using ultrasound imaging	Mid-spine line, vertebral level, interlaminar space, epidural and intrathecal spaces	1) Ultrasound increases the success rate and ease of neuraxial block performance. 2) Ultrasound usage for neuraxial procedures reduces the risk of traumatic procedures and, thus, may increase safety

Speckle noise decreases the ultrasound image contrast and makes useful clinical information harder to differentiate [12].

Zheng and his research team pioneered the application of 3D ultrasound imaging technique for scoliosis detection [13],[14]. They have developed a dedicated Scolioscan system for this ailment and it is a radiation-free, semi-automatic 3D ultrasound system. The imaging technique used in scolioscan is volume projection imaging (VPI). VPI analyses the intensity of all voxels of ultrasound volumetric data to form a coronal image [14]. It also employs the transverse process (TP) measurement method to detect the bony features in an ultrasound scan. The spinal angle measured by Scolioscan using VPI has been proven to be comparable to the earlier gold standard of Cobb angle obtained through the X-ray method [15].

While Zheng's research has been innovative, a few areas can be worked upon. For instance, this technique depended on manually annotating and measuring the scoliosis angle, which in turn depends on the examiner's judgment, expertise, and speed. As a result of this human intervention, the process can be lengthy and limited in the volumes of cases that can be handled at any given time.

As an improvement, several new measurement techniques [15],[16], [17] were developed. In [15], the spine curvature angle was obtained by deriving inflection points and by using a sixth order polynomial curve fitting method to estimate the spine curve equation. However, this is not a fully automatic process as the marking of inflection points still required human judgment. In [17], Zhou researched an automatic spine curvature measurement technique using 3D ultrasound image pre-processing with phase congruency and a newly developed two-fold threshold strategy. Though this method overcame the drawbacks of manual measurement, the computational time of this method was lengthy as most of the time was spent on computing the phase congruencies of the images.

A. Identifying suitable architecture to automate

New research was carried out to develop an alternative indicator to measure scoliosis. This indicator is called ultrasound curve angle (UCA) and was proven to be equivalent to radiographic Cobb angle. In the measurement of UCA, more lateral features of the spine are used instead of using the spinous process as an anatomical reference for angle measurement [18]. Before the calculation of the UCA, the evaluators have to first identify the exact locations of the thoracic and lumbar bony features above and below T12 level respectively and ascertain the most tilted thoracic and lumber bony feature pairs. Fig. 1. (d) shows the thoracic bony features, rib, T12 level, and lumbar bony features of a 2D ultrasound spine image.

The clear identification of thoracic bony features (TBFs) and lumbar bony features (LBFs) plays a vital role in UCA measurement. This identification process is manual, which is fully dependent upon the expertise of doctors and also

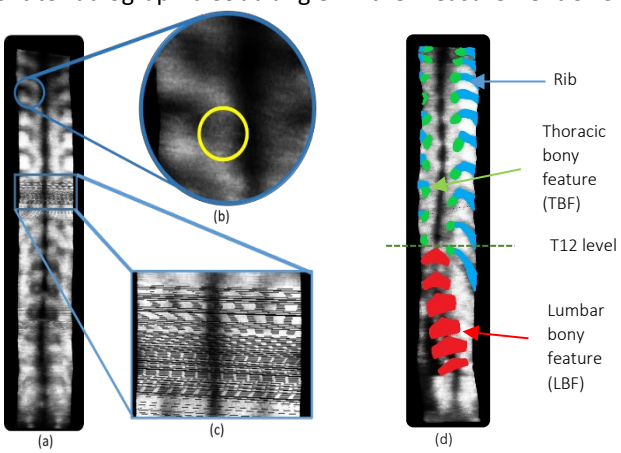

Fig 1. Ultrasound spine image (a) input image, (b) & (c) types of speckle noise (d) Various regions of interests in a spine image

time-consuming. As the first step towards automation, it is imperative that a suitable architecture should be developed which can segregate the bony features from the ultrasound image amidst the many speckles [19] and other noises. The segmentation of TBFs and LBFs could be then used to identify the most skewed regions of the spine for UCA measurement.

B. Need for a novel architecture

The locations, shapes, and sizes of the bony features are diverse (Fig. 2 (a), (b) & (c)) for different patients as are their curvature angles. The deformity of spines between two end-points of two adjacent bony features are also different (Fig. 2 (d), (e), and (f)). These variabilities of LBFs and TBFs in a noisy ultrasound image make the segmentation work more challenging.

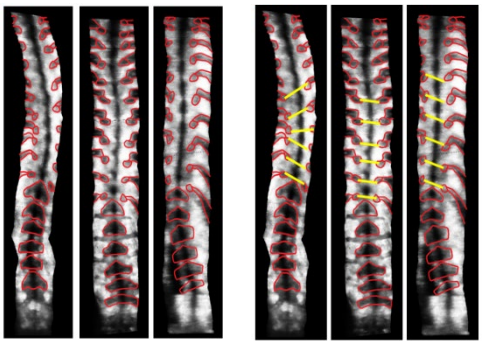

Fig 2. (a), (b) and (c) Variability in shape and size and numbers of individual Bony Features (d), (e) and (f) compounded complexities when deformities between two adjacent bony features (thoracic) are considered

C. Image processing using CNN

Usage of Convolutional Neural Network (CNN) for the analysis of medical images is a recent trend [20] and has become the most used deep learning technique in this field [21], [22], [23], [24]. Because CNN has an end-to-end characteristic, it does not require a manual design of features and has been proven excellent in image feature extraction [25]. Kokabu et al. applied a basic CNN architecture to assess the performance of a 3D depth sensor imaging system in predicting the Cobb angle [26]. However, due to the inherent process of depth scanning, the research faced a limitation in predicting the Cobb Angle with high accuracy. Also, no external validation dataset was considered, and the CNN architecture applied was computationally very expensive Also, in general, CNN's usefulness and accuracy are hampered by their need for large quantities of training data. In the case of medical images, getting hold of adequate image can often be expensive, complicated and the prerequisite of accurate annotations adds to the complexity [27].

In biomedical image segmentation, enormous success was achieved by using the U-Net architecture [28]. A U-Net is made up of two main sections: (a) Multilayer deep encoder network which helps to extract spatial features from an image, and (b) corresponding multilayer deep decoder network that up-samples the extracted feature maps to predict the final segmentation output. It utilizes the self-learning property of the convolution kernel to process the original image and delivers the classification result. Since its architecture is modular in construction, the U-Net can extract considerably complex and detailed image features just by increasing the depth or the number of layers in the architecture. Usually, the lower layers of U-Net are capable of extracting some common features of images, whereas the higher layers extract more targeted features [29]. In a basic CNN model, the spatial information may get lost during the max-pooling and transposed convolution operation. To reduce this loss of information, U-Net employs skip pathways that connect the encoder to the corresponding decoder in the same layer.

U-Net can achieve very good results with a small number of training datasets, which is very useful in the case of medical image segmentation, where a large number of training data is mostly unavailable. For instance, a two-stage U-Net was used for the segmentation and detection of breast lesions with various shapes and locations from ultrasound images with artifacts with high accuracy [30]. Also, a modified version of U-Net, Oct-U-Net was used successfully to diagnose fetal spina bifida from a 3D ultrasound image and gave better segmentation accuracy than Fully Convolutional Network [31]. Research has also been carried out on the application of U-Net in scoliosis detection such as the basic U-Net was employed to automatically segment the bony features from 2D ultrasound images for scoliosis measurement. However, as the segmentation technique was applied on sparse 2D images, the predicted segmentation accuracy was low [32].

An advanced version of U-Net, U-net with robustness to speckle and regular occlusion noise (RSN-U-net) was introduced to noise removal of spine ultrasound VPI images and segment the bony features. As the variability of shape and sizes of bony features was not addressed in that research, the segmentation accuracy was not high [33]. Recently, dual-task ultrasound transverse vertebrae segmentation network (D-TVNet), another modified U-Net along with Atrous spatial pyramid pooling (ASPP) module was introduced to segment the ultrasound spine image [34]. The research aimed to clearly distinguish boundary edges of the bony features from noisy ultrasound scan images. Though it produces an overall promising segmentation output, the architecture was insufficient to segment the bony features properly when a

larger area was occupied by noise

However, for the convolution operation, U-Net is not designed to choose the appropriate kernel size to handle the large variation in the locations, shapes, and sizes of the features [35]. Adding more layers to a conventional U-Net is not a good option as the network would become deeper and would produce redundant computation during training [36]. Moreover, the skip-connections of U-Net impose a limiting fusion scheme, forcing combinations only at the same scale feature maps of the corresponding encoders and decoders [37], [38].

A sequence of fixed multi-scale Gabor filters was used in biomedical image processing to handle the variations in an input image. In CNN, the concept of Inception architecture was introduced [39] to perform the same operation. Also, since Inception blocks utilize convolutional layers of varying kernel sizes in parallel, they can be used to assess the region of interest from diverse scales [39], [40], [41], and to combine and carry the outputs deeper into the network.

During this research, it is found that for dealing with the complexity of an ultrasound image, same-scale feature fusion skip pathways alone will not suffice. As an antidote to U-Net's restrictive same scale feature fusion problem, UNet++ aimed to improve segmentation accuracy by including (a) dense blocks [42] and (b) convolution layers between the encoder and decoder [37, 38]. Dense skip connections ensure that all prior feature maps are accumulated and transferred to the succeeding node along each skip pathway. This generates full resolution feature maps at multiple semantic levels and also the re-designed skip connections aggregates features of different semantic scales which produced highly flexible feature fusion schemes.

However, the information at the encoder side is more contextual and that at the decoder side is more localized. In U-Net++, dense skip connections originate at the encoder side and can result in speckle noises getting transferred to the decoder during the concatenation operations. It may give a suboptimal performance while handling ultrasound images. There are two major problems to be addressed in this research:

a) Choosing the appropriate kernel size for the convolution operation which can handle the large variation in locations, shapes, and sizes of the features.

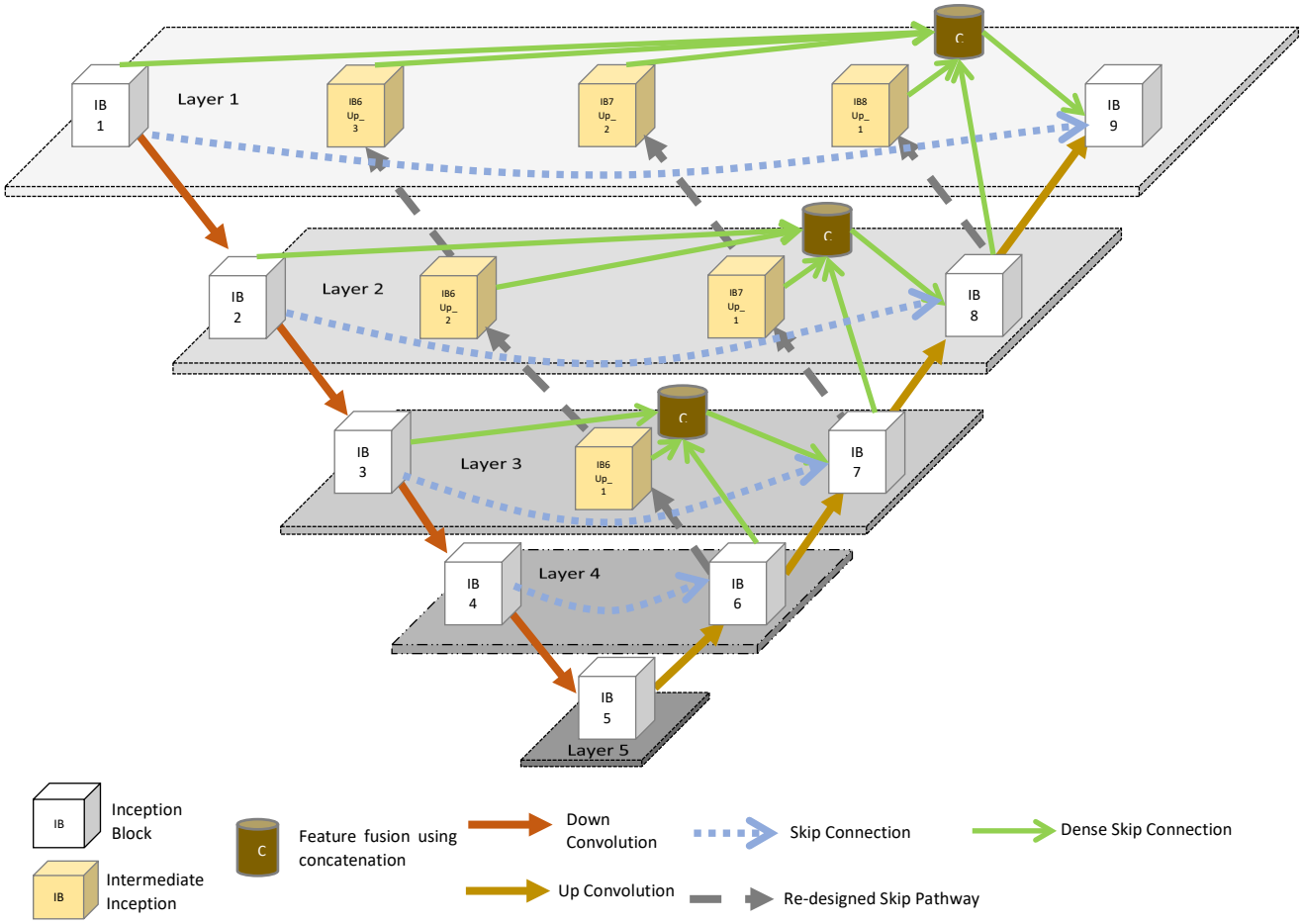

Fig 3. Overall architecture of proposed SIU-Net

b) Designing a suitable architecture that can extract semantically rich features and fuse multi-scale features for a better segmentation output.

In this research, a novel deep learning architecture, Skip-Inception U-Net or SIU-Net, is developed to suitably segment the TBFs and LBFs in the ultrasound spine image dataset. The standard U-Net is adopted as the main network architecture, and the simple convolutional layers are replaced with Inception blocks [40]. The encoders-decoders are bridged using newly designed decoder side skip-pathways [38]. The segmentation result from the proposed architecture is compared with the basic U-Net, UNet++, and MultiResUNet [35] models. The result shows that the new model outperforms the aforementioned models.

II. METHODOLOGY

The architecture of the proposed Skip-Inception U-Net or SIU-Net is shown in Fig 3. This network model employs the basic U-Net structure as the base framework and contains the improvised Inception blocks, the re-designed Dense-skip connection feature fusion using concatenation and the Down-sample path, and the Up-sample path.

A. Inception Block

To solve the issue of choosing an appropriate kernel size to handle large variability in the spine image dataset, the concept of Inception block (IB) is adopted to develop a high-performance segmentation model [39]. A modified IB (Fig. 4) of Inception V2 architecture [40] is introduced to replace the traditional convolutional layers of basic U-Net. There are two advantages of IB: (a) it increases the depth and width of the model without any increase of computational requirement and (b) it allows the flexibility of using multiple filter sizes within the same level [43].

In the main U-Net architecture, a sequence of two 3×3 convolutional layers was used after each pooling layer and transposed convolutional layer. On the other hand, a basic Inception block involved the use of a 5×5 filter, which required high computation. According to [40], factorizing one 5×5 convolution to two 3×3 convolution operations can improve the computational speed since one 5×5 convolution is 2.78 times more time consuming than one 3×3 convolution.

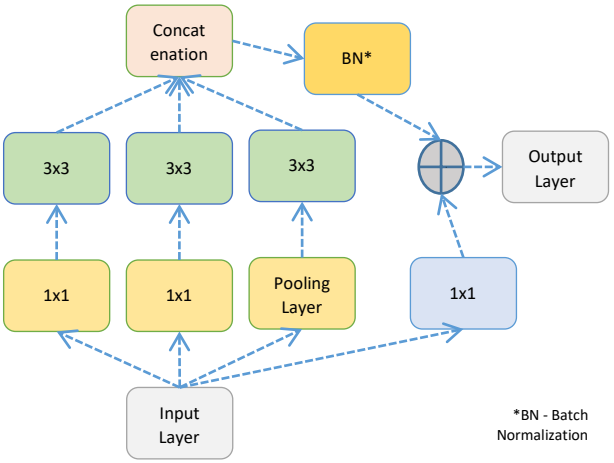

Fig 4. Inception Block in proposed architecture.

In SIU-Net, the selection of the filter sizes of the modified IB is optimized to 3×3 and 1×1 to make it computationally efficient. The modification enables better aggregation of feature maps from different branches of kernels of different sizes and makes the network wider and capable of learning more features [44]. Also, the residual connection makes the learning easier since a residual IB learns a function with reference to the input feature maps, instead of learning an unreferenced function [41]. Unlike the original Inception V2 architecture, each 1×1 convolutional layer is followed by a 3×3 convolution layer. The concatenated 3×3 layers are then followed by a batch normalization (BN) layer that avoids gradient vanishing while retaining the convolutional layers.

In the modified Inception block, the technique is adapted from [35] to control the number of filters of the convolution layers inside the block. A parameter P is assigned to control the number of filters as given by equation -

$$P = \lambda \times F \qquad (1)$$

where, F is the number of filters in the corresponding layers similar to the basic U-Net, and λ is a scaler coefficient. The number of filters is set to F = [32, 64, 128, 256, 512] along with the layers respectively, and λ is chosen as 1.67 to ensure that the model structure is similar to the basic U-Net. The number of filters is set to $(\frac{P}{6}, \frac{P}{3}, \frac{P}{2})$ in the three corresponding convolutional layers for extracting multiscale features.

B. Dense-skip connection

A multi-scale feature fusion scheme impacts segmentation accuracy more than a single-scale [38]. In SIU-Net, the output of the previous IB of the decoder of the same dense block is merged with the corresponding up-sampled output

of the lower dense block through dense skip connections (DSC) (Fig. 3). Through a dense convolution operation, each node in a decoder is presented with a final aggregated fused feature map containing: a) the feature from the previous decoder, b) intermediate block combined feature maps and c) the same-scale feature from the corresponding encoder.

Let,($f^{m,n}$) denote the output of node $F^{m,n}$ where *m* is the down-sampling layer along the encoder side and *n* is the convolution layer of the dense block along with the skip connection. The whole feature fusion scheme is shown in Fig 5. The fused feature maps $f^{m,n}$ is calculated as:

$$f^{m,n} = \begin{cases} \dot{\mathrm{E}}\left(\mathrm{D}\left(f^{m-1,n}\right)\right), & n = 0 \\ \dot{\mathrm{E}}\left(C\left[\left(f^{m,p}\right)_{p=0}^{n}, U\left(f^{m+1,n-1}\right)\right]\right) & n > 0 \end{cases} \qquad (2)$$

Where $\dot{\mathrm{E}}(.)$ denotes a convolution operation followed by the activation function, D (.) and U (.) denotes a down-sampling and an up-sampling layer respectively and C [ ] denotes the concatenation operation. Nodes at level n=0 receive only one input from the previous layer of the encoder, the nodes at level n=1 get features from the same level encoder, one encoder sub-network, and the previous layer decoder. Nodes at level n>1 get n+2 features, where the two features are from the same layer encoder and previous layer decoder respectively and the rest of the n features are from n encoder sub-networks of the same skip-connection. Encoder sub-networks are the up-sampled output from the lower level skip-connection.

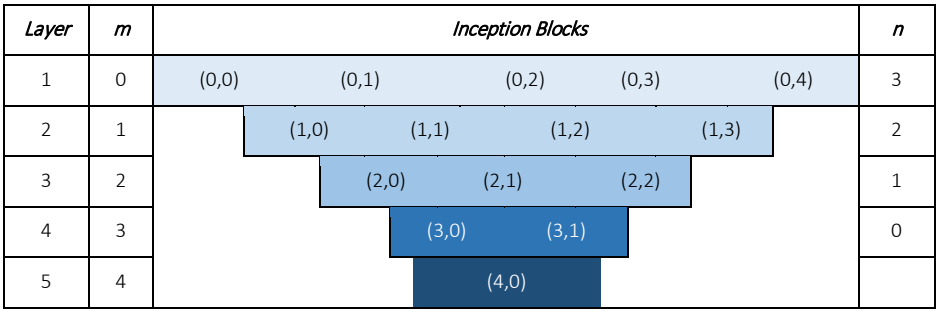

Fig 5. Feature fusion scheme using the dense convolution operation

Therefore, the top layer of the proposed network (Fig. 3) is:

$$IB9 = \dot{\mathrm{E}}(C[IB1, IB6up_3, IB7up_2, IB8up_1, U(IB8)]) \qquad (3)$$

Where $\dot{\mathrm{E}}(.)$ indicates a convolution operation followed by the activation function, *C* [ ] indicates the concatenation operation and *U(.)* indicates up-sampling. Other layers follow similarly.

By this, the advantage offered by the skip pathway is retained. Since the shallower layers of the decoder side are used to extract the information, more localized information is extracted by SIU-Net. Also, these DSCs generate full resolution feature maps at multiple semantic levels and help to improve segmentation accuracy and gradient flow.

C. Ablation Study:

During the initial stages of research to find the best segmentation method for segmenting the TBFs and LBFs in the ultrasound spine images, several models were designed and continuously improved from the basic U-Net structure. Notably, three models are highlighted below.

In the first model, the improvised IB (Fig. 4) is employed within the basic U-Net architecture while the down-convolution, up-convolution, and skip pathways are unchanged. This network is named as IU model (Inception+ U-Net). The output is not satisfactory as the issue of the restrictive single-scale feature aggregation is not addressed.

In the second model, the idea of Residual path [35] of MultiResUNet is deployed by incorporating some convolution layers with residual connections along the traditional skip path. This is done to add some additional non-linear transformations on feature propagation from the corresponding encoder to the decoder. This new model is similar to the normal MultiResUNet except that the conventional IB is replaced with the modified Inception block. This network is named as IRs model (Inception + Res path). However, the performance of this model is found unsatisfactory because the residual paths do not extract features from deep layers and are inadequate to handle noisy ultrasound images.

In the third model, along with the improvised IB, the encoder side skip-pathways (similar to UNet++) are introduced. It

is named as ISP model (Inception + encoder skip path). Though this model gives better output than the previous two models, it is found that the absence of the skip pathways in the decoder side results in a lack of richness of information especially during edge detection of bony features amidst image noises.

III. Experimental Setup

A. Dataset

Among all available imaging techniques, a 3D ultrasound volume projection image is the best suitable to visualize spine anatomy [9]. The result is very satisfactory when compared with the conventional radiographic Cobb angle method. The Scolioscan system (Model SCN801, Telefield Medical Imaging Ltd), developed in Hong Kong, is used to generate 3D volume projection image (VPI) using the conventional 3D ultrasound imaging technique.

The approval for conducting the experiments involving human subjects was given by the Institutional Review Board. The patients provided informed consent for including them in this study as required, and the work complies with the Declaration of Helsinki. A total of 109 images, collected from 109 patients (82 females and 27 males) with an average age of 15.6 ± 2.7 years, are used retrospectively. The patients suffer from different degrees of spine deformity, which puts further challenges to the generalization of our proposed framework.

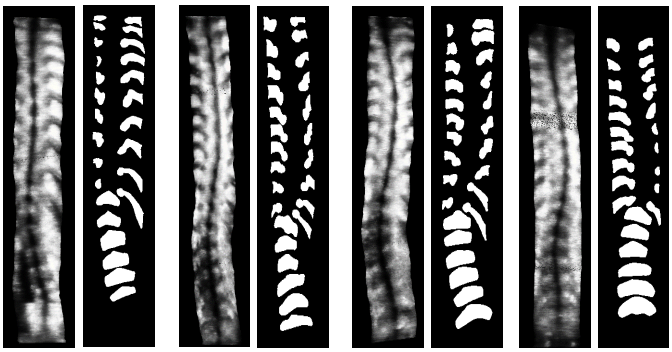

Fig 6: Four Image sets with each having input image with respective truth mask

The input data set (Fig. 6) consists of:

a) Input Data set: Nine 2D vertebral anatomical images were extracted from one volumetric image at specific depths[15]. For future processing, one 2D image is first resized to 2574×640 pixel and then saved in '.png' format.
b) The expert suggested Truth Mask (TM): Experts from Hong Kong Polytechnic University have created a truth mask for each input data image and have made it available for this research.

The annotation of the truth mask is done based on some important features. Firstly, the lumbar bony features (LBF) shall be six in number. Secondly, if the last pair of T12 level (Fig. 1(d)) is not clearly visible, it is annotated according to the experience of the experts.

The input data sets of 109 2D spine ultrasound images are randomly categorized into two sets, one having 79 images as training set and other having 30 images as test set. Further, data augmentation is done to increase the size of the training dataset by randomly flipping and rotating the available images.

Each input ultrasound image, of size 2574 × 640 pixels, is resized to 256 × 64 pixels through image pre-processing and maintaining the aspect ratio of the original image.

B. Implementation Details

Software: Spyder, Anaconda.

Libraries: Keras with Tensorflow backend [45].

Machine: GPU laptop having NVIDIA GeForce RTX 2060.

The mode of operation of any semantic segmentation algorithm is to analyze each pixel and to predict whether they represent a point of interest, or are merely a part of the background. This means that the task can also be looked as a pixel-wise binary classification problem where the objective of the segmentation algorithm would be to minimize the binary cross-entropy loss function.

For an image *X*, let the corresponding Truth Mask (TM) be *Y*, and the predicted segmentation output be *Y′*. For a pixel *px*, the TM value is y_{px} and the network predicted output is y'_{px}, The binary cross-entropy loss for that image is defined as:

$$Cross\ Entropy\ (X,Y,Y') = \sum_{px\epsilon X} [\![-\left(y_{px}\right]\!] \log\left(y'_{px}\right) + \left(1-y_{px}\right)\log\left(1-y'_{px}\right)\Big) \quad (4)$$

For a batch containing n images, the loss function J becomes,

$$J = \frac{1}{n}\sum_{i=1}^{n} Cross\ Entropy\ (X_i, Y_i, Y'_i) \tag{5}$$

The binary cross-entropy loss is minimized, and the model is trained using Adam optimizer [46]. Adam optimizer adaptively computes different learning rates for different parameters from estimates of the first and second moments of the gradients. All the models are trained up to 150 epochs since 150 epochs are found to be the saturation point for model accuracy, and no further improvement is observed beyond this point. Finally, for validating the reliability of the model, 5-fold cross-validation of the dataset is done.

C. Evaluation Metrics

For quantitative performance evaluation, three very popular evaluation indices are employed – Jaccard similarity, Dice Coefficient, and Euclidean distance. The Jaccard similarity (JS) [47] and DICE coefficient (DC) [48] and Histogram Euclidean distance (d) are given by,

$$Jaccard\ Similarity = \frac{J \cap \hat{J}}{J \cup \hat{J}} \tag{6}$$

$$Dice\ coefficient = 2\left(\frac{J \cap \hat{J}}{J + \hat{J}}\right) \tag{7}$$

$$Histogram\ Euclidean\ distance, d = \sqrt{\sum_i (hist(i1) - hist(i2))^2} \tag{8}$$

$\hat{J}$ is the predicted segmentation output from the method to be evaluated, and J is the expert suggested Truth Mask. These J contours are references for further segmentation analysis [49]. $hist(i2)$ is the histogram of the predicted image and $hist(i1)$ is the histogram of the corresponding truth mask.

IV. Results

A thorough analysis of the proposed model, SIU-Net, is undertaken and compared with basic U-Net, UNet++, and MultiResUNet. Also, the performances of the intermediary models, i.e. IU model (Inception+ U-Net), IRs model (Inception + Res path), and ISP model (Inception + encoder skip path), are analyzed. All the architectures have the same settings as illustrated in section III.B.

TABLE 2
QUANTITATIVE EVALUATION

Method	Avg. Jaccard Index (Std. Dev.)	Avg. Dice Score (Std. Dev.)
U-Net	0.709 (0.034)	0.817 (0.037)
MultiResUNet	0.732 (0.034)	0.853 (0.029)
UNet++	0.748 (0.035)	0.861 (0.034)
Incep. + U-net	0.711 (0.035)	0.846 (0.033)
Incep. + Res path	0.734 (0.035)	0.855 (0.029)
Incep. + Encoder Skip Path	0.757 (0.036)	0.876 (0.026)
Proposed SIU Net	**0.781 (0.033)**	**0.883 (0.025)**

A. SIU-Net outperforms U-Net, UNet++, and MultiResUNet in the segmentation of ultrasound spine image dataset:

Table 2 details the segmentation performance (JS and DC) of the SIU-Net, UNet++, MultiResUNet, U-Net, and other

three intermediary models i.e. IU, IRs, and ISP. It can be observed that each intermediary model (i.e. IU, IRs, and ISP) has a significant improvement on segmentation performance over their corresponding basic methods. This is because an IB, by its nature of construction, is more adept in extracting features from different locations, shapes, and sizes than conventional convolution layers. Fig 7 illustrates the performances of all the models in terms of the Jaccard index, Dice index, and Histogram Euclidean distance. The evaluation result shows that SIU-Net is deeper and more capable of learning

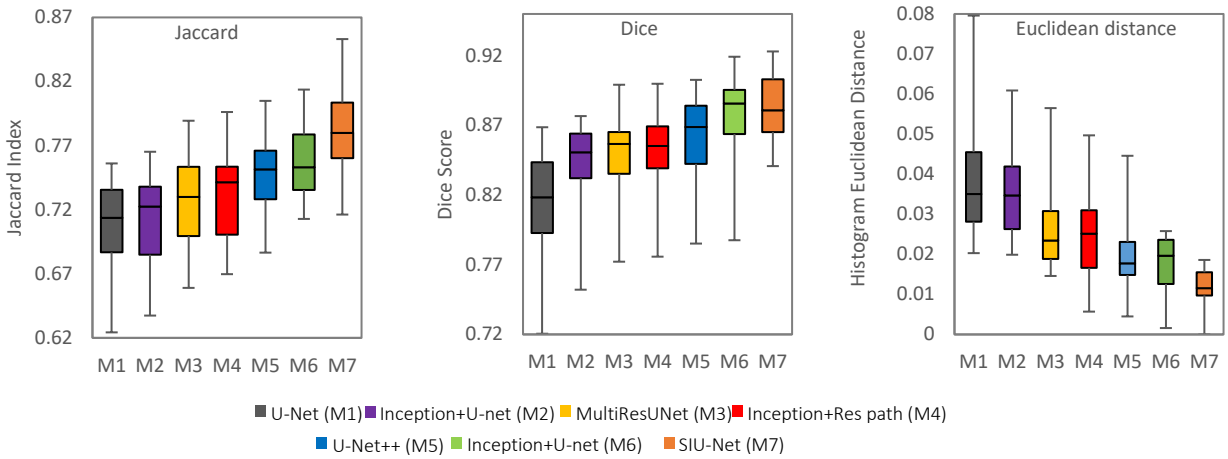

Fig 7: Performance comparison of models using (a) Jaccard Index, (b) Dice Index and (c) Histogram Euclidean distance

features from datasets and achieves a better performance in terms of TBF and LBF segmentation.

B. *SIU-Net performs best for bony feature edge detection:*

The main aim of this research is to appropriately locate LBFs and TBFs from the ultrasound spine images with a high degree of clarity for accurate scoliosis detection. This requires the segmentation of each bony feature with a proper edge boundary and location. Because of speckle noise, ultrasound images suffer from the lack of clear boundaries which complicates the segmentation process. Though U-Net and MultiResUNet can segment the bony features from the input

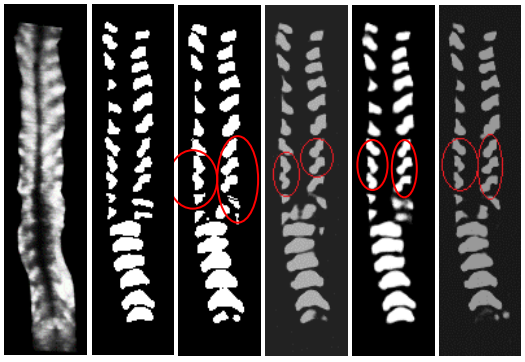

Fig 8.a. SIU-Net is able to distinguish TBF pairs

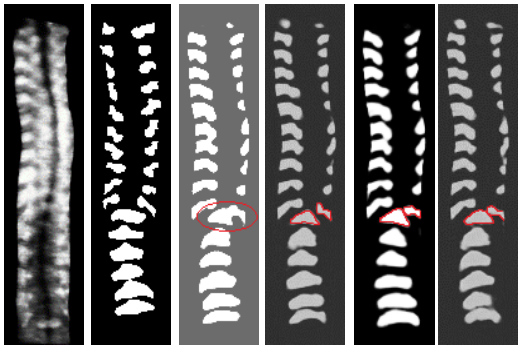

Fig 8.b. SIU-Net is able to distinguish TBF and LBF

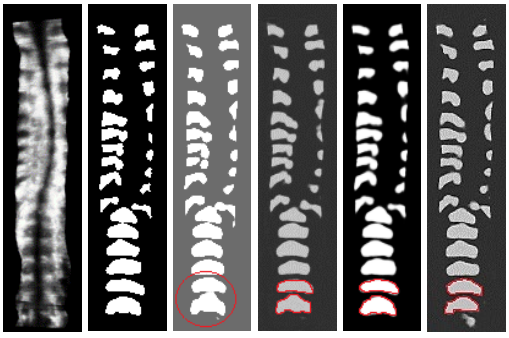

Fig 9.a. SIU-Net is able to segment LBF

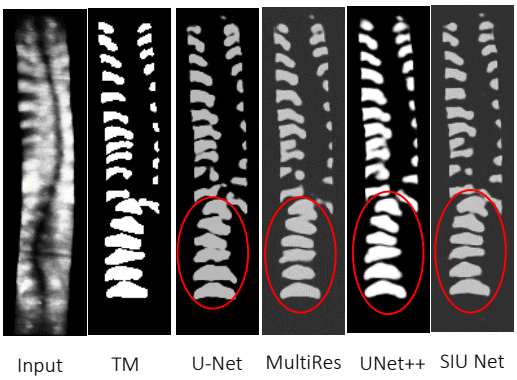

Fig 9.b. SIU-Net is able to identify six individual LBFs

images, UNet++ and SIU-Net perform better to clearly distinguish the edge boundaries (Fig. 8.a). Further detailing suggests that UNet++ performs better than MultiResUNet, but it cannot outperform SIU-Net. The overall performance of a particular image is explained in the Fig. 8.a. The 8th, 9th and 10th TBFs pair's edge boundaries are not distinguishable using U-Net (JS: 0.7198, DC: 0.8201) and MultiResUNet (JS: 0.7241, DC: 0.8367), but it is prominent using UNet++ (JS: 0.7322, DC: 0.8484) and SIU-Net (JS: 0.7415, DC: 0.8503).

For more challenging images (Fig. 8.b), the 10th right-side TBF and 1st LBF are indistinguishable in U-Net (JS: 0.7313, DC: 0.8393) but clearly distinguishable in MultiResUNet (JS: 0.751, DC: 0.8578), UNet++ (JS: 0.7649, DC: 0.8601) and SIU-Net (JS: 0.7723, DC: 0.8679).

C. SIU-Net performs the best identification of LBF:

There are six lumbar bony features visible in an ultrasound spine image. Though basic U-Net and MultiResUNet give good Jaccard and Dice value, in some images, the LBFs are not segmented. In Fig. 9.a, the two bottom LBFs are indistinguishable in U-Net segmentation (JS: 0.7677, DC: 0.8542), but they are segmented in MultiResUNet (JS: 0.7801, DC: 0.8708), UNet++ (JS: 0.7953, DC: 0.8839) and SIU-Net (JS: 0.8001, DC: 0.8973). With closer observation, it can be seen that in SIU-Net, the edge boundary is better than the other two.

Again from Fig. 9.b, it is clear that, in case of U-Net (JS: 0.7074, DC: 0.8158) and MultiResUNet (JS: 0.7179, DC: 0.82), many LBFs appear to be conjoined, whereas outputs of UNet++ (JS: 0.7265, DC: 0.834) and SIU-Net (JS: 0.7319, DC: 0.8418) are able to distinguish all the 6 LBFs properly.

D. SIU-Net is most capable of identifying TBF:

Speckle noise poses considerable challenges, especially during ide ntifying TBFs, which are more in number and smaller in size than LBFs. In Fig. 10.a, as shown in the marked red area, the right TBFs are quite vague in the input image. While the other two methods (U-Net (JS: 0.6633, DC: 0.7976), UNet++ (JS: 0.6598, DC: 0.7936)) are able to identify one or two of those TBFs, MultiResUNet (JS: 0.6583, DC: 0.792) and SIU-Net (JS: 0.6494, DC: 0.7874) are better in differentiating all of them. Overall, SIU-Net performs better than MultiResUNet as T12 level is fragmented in the case of the latter (shown in yellow marked area). In fact, SIU-Net performs better than the manual segmentation process as those three TBFs are

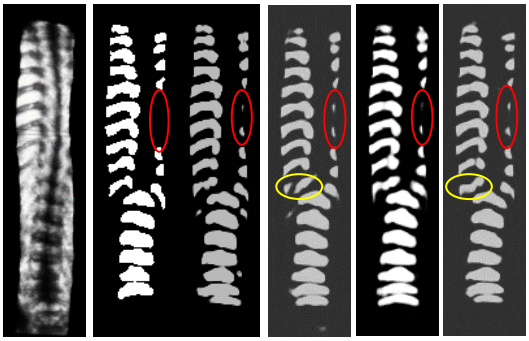

Fig 10.a. SIU-Net is able to identify TBFs that are invisible in TM

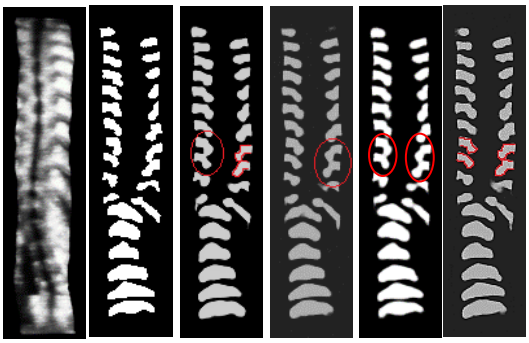

Fig 10.b. SIU-Net is able to detect the edges of TBFs that are indistinguishable in other two methods

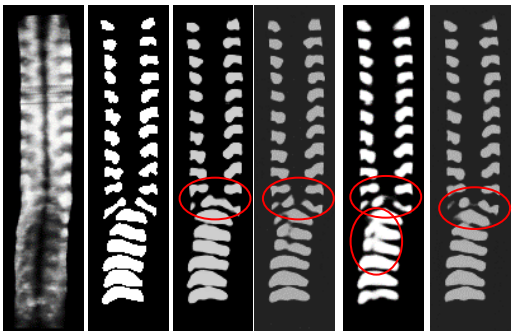

Fig 11.a. SIU-Net outperforms MultiResUNet in middle bony feature segmentation

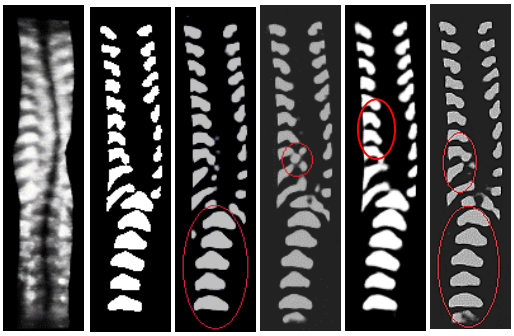

Fig 11.b. SIU-Net is able to distinguish both TBFs and LBFs clearly

missing even in the truth mask.

In Fig. 10.b, the 8th and 9th TBFs are conjoined in U-Net (left side) (JS: 0.7812, DC: 0.8623), MultiResUNet (right 7th and 8th TBFs) (JS: 0.808, DC: 0.8801) and UNet++ (both sides) (JS: 0.8113, DC: 0.8867). SIU-Net (JS: 0.831, DC: 0.9056) can define both the pairs clearly and also gives the highest Dice and Jaccard score. It can be concluded that SIU-Net performs better identification of TBF pairs than the other three methods.

E. SIU-Net is more consistent than MultiResUNet and UNet++:

From the previous cases, it is ascertained that, while all of the three models (UNet++, MultiResUNet, and SIU-Net) outperform U-Net, the performance of SIU-Net is the best. Upon closer scrutiny of the last pair of TBFs in Fig. 11.a, it is found that they are indiscernible in U-Net (JS: 0.7703, DC: 0.8588). But both are identified in MultiResUNet (JS: 0.7893, DC: 0.8607), UNet++ (JS: 0.7948, DC: 0.8798) and SIU-Net (JS: 0.8203, DC: 0.8974). However, the performances of the other methods with respect to the overall clarity of LBFs and TBFs are inconsistent.

Also, in Fig. 11.b, U-Net (JS: 0.711, DC: 0.8375) is unable to identify 6 LBFs but able to segment all the TBFs clearly,

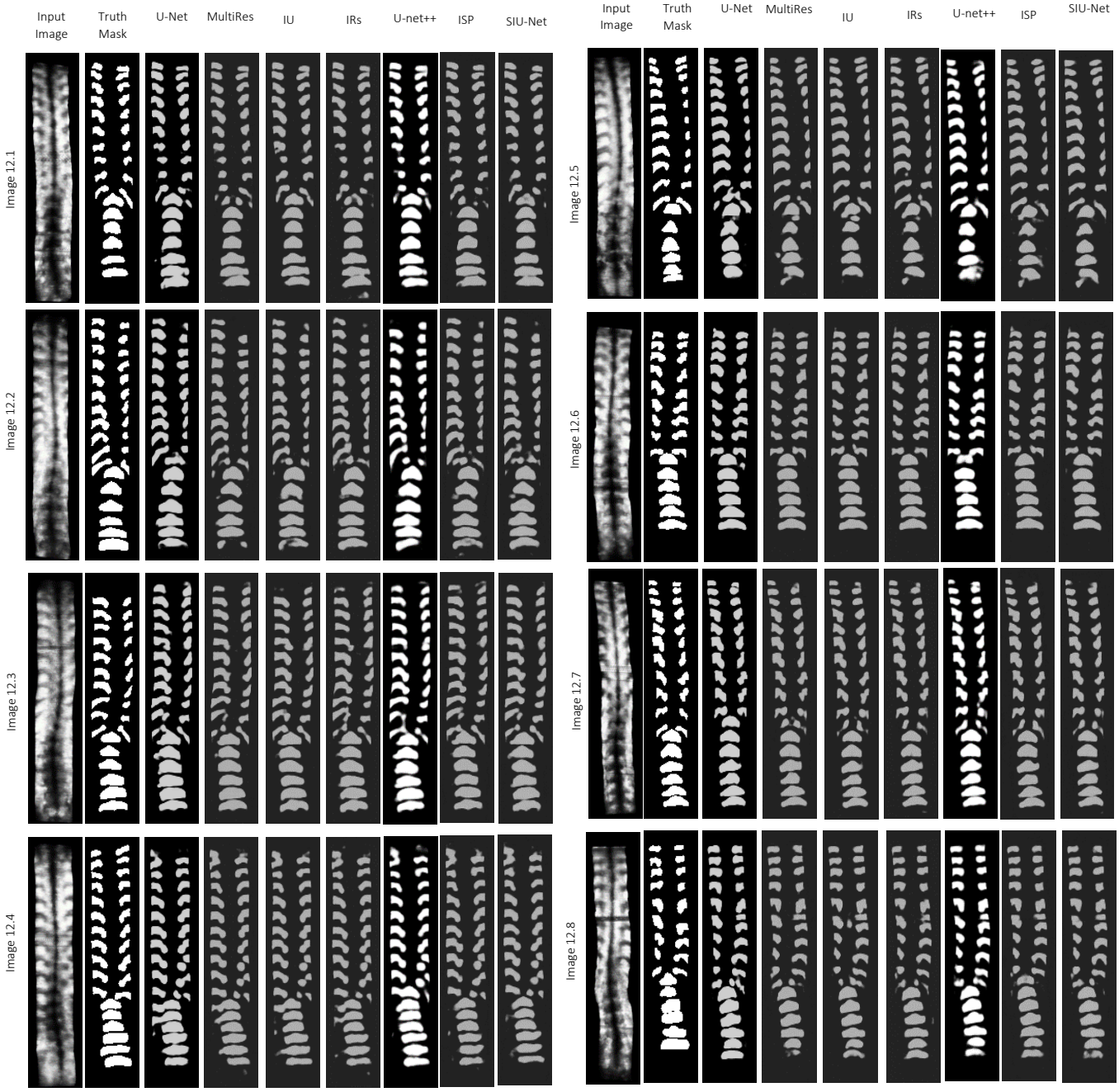

Fig 12. Qualitative comparison of all models using ultrasound spine image

whereas MultiResUNet (JS: 0.7509, DC: 0.8551) and UNet++ (JS: 0.7698, DC: 0.8698) identify 6 LBFs but fail to identify the 8th and 9th left TBFs. Unlike the other three methods, SIU-Net (JS: 0.7813, DC: 0.8916) identifies all the 6 LBFs clearly and segments all TBF pairs. However, the SIU-Net outperforms all the other models in the successful segmentation of both the TBFs and LBFs for its unique architecture. Hence, it can be concluded that SIU-Net gives the most consistent performance when compared to basic U-Net, UNet++, and MultiResUNet.

F. Qualitative and Quantitative comparison of all models:

Fig. 12 shows a few more qualitative visual comparisons of bony feature segmentation results of the ultrasound spine image using SIU-Net and other methods: basic U-Net [28], UNet++ [38], MultiResUNet [35], IU, IRs model, and ISP model. Table 3 summarizes the Jaccard and Dice values of each image. In each of these cases, SIU-Net also qualitatively outperforms all the other segmentation methods.

TABLE 3
QUANTITATIVE EVALUATION OF INDIVIDUAL IMAGE

Image No	U-Net		MultiResUNet		Inception + U-Net		Inception + Res path		Unet++		Incep. + Encoder Skip Path		SIU-Net	
	Jac.	Dice	Jac.	Dice	Jac.	Dice	Jac.	Dice	Jac.	Dice	Jac.	Dice	Jac.	Dice
12.1	0.702	0.825	0.689	0.836	0.679	0.829	0.685	0.838	0.729	0.847	0.741	0.850	**0.758**	**0.853**
12.2	0.713	0.832	0.728	0.839	0.713	0.833	0.729	0.839	0.732	0.842	0.759	0.847	**0.763**	**0.851**
12.3	0.707	0.828	0.731	0.844	0.722	0.830	0.738	0.845	0.747	0.861	0.758	0.869	**0.767**	**0.873**
12.4	0.717	0.840	0.729	0.849	0.719	0.841	0.730	0.849	0.738	0.853	0.751	0.859	**0.759**	**0.863**
12.5	0.735	0.847	0.760	0.863	0.750	0.849	0.763	0.860	0.789	0.871	0.801	0.879	**0.820**	**0.883**
12.6	0.808	0.880	0.809	0.895	0.808	0.885	0.81	0.896	0.820	0.903	0.828	0.911	**0.841**	**0.919**
12.7	0.760	0.863	0.775	0.873	0.764	0.867	0.77	0.870	0.786	0.887	0.791	0.893	**0.803**	**0.901**
12.8	0.636	0.778	0.656	0.792	0.638	0.779	0.657	0.791	0.678	0.823	0.692	0.844	**0.703**	**0.853**

G. Comparison of accuracy of all models:

The goal of a segmentation model is to accurately detect all the bony features which are present in their respective truth masks. A quantitative study is done to assess the accuracy in which the cases of the conjoint and broken features in an output segmentation image are considered to be failure cases. Also in this study, it is found that, in general, the chances of conjuncts are particularly high in the case of LBFs and the overall accuracy of detecting LBFs is lower than that of the TBFs.

Fig. 13 depicts the comparison of all models in terms of percentage of images where all the bony features as well as individual thoracic and lumber bony features are detected clearly. From the figure, it is can be concluded that SIU-Net is the most accurate model, amongst the other models, to detect the bony features.

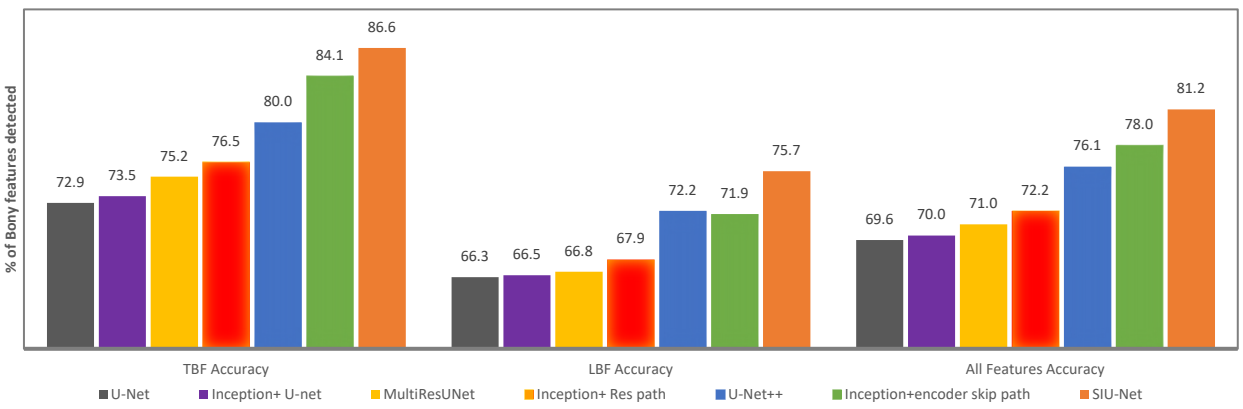

Fig 13: Proportion (%) of Images where segmentation outputs are able to detect all Bony features from respective Truth masks

V. DISCUSSION

The quality of input images significantly impacts the quality of segmentation output. Among the imaging modalities used for scoliosis assessment, X-ray, MRI, and CT produce superior quality images while the non-radiating imaging modalities such as ultrasound are susceptible to noise and low contrast. U-Net is one of the most popular segmentation networks in deep learning. In previous work, U-Net performed very well in spine segmentation from X-ray images [28]. In a fully automated measurement of sagittal spinopelvic balance, U-net performed well in detecting anatomical landmark probability maps on the sacral endplate from sagittal X-ray images [50] because the X-ray image is more apparent than the ultrasound image and, therefore, access to segment. There are inherent limitations of ultrasound imaging compared to X-ray, due to which ultrasound cannot always compete with X-ray imaging. Acoustic shadowing hides all bone surfaces that are deeper than the posterior surface of vertebrae and ribs. Therefore, basic U-Net is unable to segment all relevant features and provide missing bony features in segmentation output consistently. Table 4 compares the Dice score of multiple segmentation methods, both automatic and manual, using X-Ray [51], CT, and MRI [52] with ultrasound. In the case of CT, the authors indicated that their method were not designed to tackle low-dose CT images and images with implants and only the medium to high-quality CT images were selected (dataset 3) for further segmentation [53]. The proposed method, SIU-Net, explicitly designed to manage the inherent drawbacks of ultrasound images, produces results comparable (by 93-95%) to other methods which used higher quality input images as input. Additionally, from the respective publications, it can be inferred that when compared to respective baseline U-Net outputs, SIU-Net gives the highest improvement (0.883 vs 0.817 – 8%) when compared to Residual U-Net (0.951 Vs 0.9410 - 1%) and MANet (0.925 Vs 0.9008 – 3%).

Ultrasound imaging modalities that are non-radiating, safe, portable, and capable of real-time operation are explored for scoliosis assessment [54]. Table 5 illustrates some popular ultrasound imaging techniques and the correlation of their respective measured scoliosis curvature angles to the radiographic Cobb angle method [55], [56], [57], [58]. Huang et al. introduced a double-sweep 2.5-dimensional extended field-of-view (EFOV) method that showed a good correlation of 0.993 [56]. However, in their research, the authors deduced that the surfaces of the transverse process may not be consistently visualized when scanning patients with scoliosis. The reconstruction of the panorama images was more time-consuming than the original image-cutting method. Chen et al. worked on 3D ultrasound with a fast reconstruction algorithm and found that it is impossible to apply the conventional slicing technique to visualize the hidden bony features of each layer from reconstructed 3D images [57]. Zheng and his team developed a dedicated system called Scolioscan

TABLE 4
COMPARISON OF SEGMENTATION PERFORMANCE OF VARIOUS ARCHITECTURES IN SPINE IMAGES USING VARIOUS IMAGING MODALITIES

Author	Objective	Imaging Modality	Method	Method	Description of dataset used	Avg. Dice Score
Horng et al. 2019 [51]	Segmentation of vertebrae for scoliosis assessment	X-ray	Residual U-Net	Automatic	595 vertebra images; Five-fold cross validation (each fold images augmented to 1000, 10% validation images)	0.951 ± 0.03
Khandelwal et al. 2021 [53]	Segmentation of entire spine and individual vertebra to aid surgical planning	Computed Tomography (CT)	Region-based segmentation of spine	Manual	Dataset 1 (Lumbar vertebrae; 50 vertebrae extracted from 10 images)	0.924 ± 0.013
				Manual	Dataset 2 (Thoracular-Lumbar Vertebrae; 120 thoracic and 50 lumbar vertebrae across 10 subjects)	0.949 ± 0.022
				Manual	Dataset 4 (Cervical, thoracic and lumbar region; 43 medium-to-high dose CT images)	0.849 ± 0.753
			Shape-prior based flows for individual vetrebra segmentation	Automatic	Dataset 3: (Lumbar vertebrae; lumbar region of 30 patients, no. of slice range 55-200) (25 for training, 5 for evaluation), six-fold cross-validation	0.838 ± 0.031
Li et al. 2021 [52]	Segmentation of spine T2-weighted images to assess lumbar spinal stenosis	Magnetic Resonance Imaging (MRI)	Dual branch multi-scale attention network (MANet)	Automatic	1080 images of 120 patients (70% training, 20% validation, 10% testing)	0.925
Proposed Method	Segmentation of thoracic and lumbar bony features to assess Scoliosis	Ultrasound	Skip-Inception U-Net (SIU-Net)	Automatic	109 images from 109 patients (79 training, 30 testing), Five-fold cross validation	**0.883 ± 0.024**

which employs 3D ultrasound scanning using the volume projection imaging (VPI) technique [13]. A coronal view of the spine similar to a posterior-anterior radiograph can be obtained from the reconstructed volume data using a simple re-slicing or volume rendering technique. One 3D ultrasound image generated is then split into nine 2D VPI images of different depths [17]. The re-slicing process helps to make the hidden bony features inside the spine volume more visible.

TABLE 5
COMPARISON OF SCOLIOSIS CURVATURE ANGLES MEASURED USING VARIOUS TYPES OF ULTRASOUND IMAGES WITH TRADITIONAL COBB ANGLE

Author	Dimensions	Type of ultrasound images	Correlation (R^2) with traditional radiographic cobb angle
Huang et al. 2017 [55]	2.5D	2.5D extended field-of-view (EFOV)	-
Huang et al. 2019 [56]	2.5D	Double-sweep 2.5-dimensional extended field-of-view (EFOV)	0.993
Chen et al. 2021 [57]	3D US with fast reconstruction algorithm	Fast Dot-Projection algorithm- Voxel-based Nearest Neighbour (FDP-VNN)	Rater 1: 0.95 Rater 2: 0.90
		Fast Dot-Projection algorithm- Multiple Plane Interpolation (FDP-MPI4)	Rater 1:0.97 Rater 2: 0.97
Brink et al. 2018 [58]	3D	Volume Projection Imaging- Spinous Process (VPI-SP)	Automatic: 0.991 (Thoracic), 0.983 (lumbar) Manual: 0.987 (Thoracic), 0.970 (lumbar)
	3D	Volume Projection Imaging- Transverse Process (VPI-TP)	Manual: 0.992 (Thoracic), 0.985 (lumbar)

This makes images from Scolioscan better suited for automatic scoliosis curvature assessment. Further, using VPI images, Brink et al. demonstrated high correlations of scoliosis curvature angle with the traditional Cobb angle method [58], for both thoracic and lumbar regions as shown in table 5.

Table 6 summarizes the research work on all the three indices that are used with ultrasound imaging for scoliosis

TABLE 6
COMPARISION OF PERFORMANCE OF VARIOUS ULTRASOUND SCOLIOSIS MEASUREMENT INDICES WITH RADIOGRAPHIC COBB ANGLE

Author	Ultrasound System	No. of patients	Method	Region of Interest (ROI)	Outcome vs Radiographic Cobb Angle	
					MAD	Correlation (R2)
ULTRASOUND COBB ANGLE						
Zheng et al. 2016 [59]	SonixTABLET	65	Manual (Blinded & AOR)	Coronal Curvature	Rater 1: 4.9°± 3.8° Rater 2: 4.6°± 3.8° (Blinded) Rater 1: 2.8°± 2.2° Rater 2: 2.7°± 1.9° (AOR)	Rater 1: 0.58 Rater 2: 0.58 (Blinded) Rater 1: 0.84 Rater 2: 0.87 (AOR)
Young et al. 2015 [60]	SonixTABLET US system	20	Manual (centre of lamina (COL) method)	Coronal Curvature	Rater 1: 2.6°± 2.0° Rater 2: 4.1°± 2.6° Rater 3: 3.8°± 3.3° Rater 4: 3.7°± 3.5°	-
SPINOUS PROCESS ANGLE (SPA)						
Zeng et al. 2019 [61]	SonixTABLET	50	Semi-Automatic (Using gradient vector flow (GVF) snake model)	Spinous process	Rater 1: 5.8° Rater 2: 6.6°	Rater 1: 0.75 Rater 2: 0.73
Zeng et al. 2021 [62]	SonixONE	92	Semi-Automatic (Using Stacked Hourglass Network)	vertebral spinous process (SP) & laminae	Rater 1: 5.7°± 4.5° Rater 2: 6.1°± 4.8°	Rater 1: 0.80 Rater 2: 0.75
S. Reuver et al. 2021 [63]	Scolioscan	70	Manual	Spinous Process	Thoracic: 6.5°± 3.9° Lumbar: 7.3°± 4.7°	Thoracic: 0.968 Lumbar: 0.923
Banerjee et al. 2020 [64]	Scolioscan	109	Automatic (U-Net segmentation)	Spinous Process	-	-
Brink et al. 2018 [58]	Scolioscan	33	Manual	Spinous column Profile	4.5° ± 3.1°	Thoracic: 0.987 Lumbar: 0.970
Brink et al. 2018 [58]	Scolioscan	33	Automatic	Spinous column Profile	4.9°± 3.2°	Thoracic: 0.991 Lumbar: 0.983
Zhou et al. 2017 [17]	Scolioscan	99	Automatic (using Phase congruency method)	Spinous column Profile	-	0.83
Zheng et al. 2016 [54]	Scolioscan	49	Manual (VPI-SP method)	Spinous column Profile	-	Thoracic: 0.784 Lumbar: 0.727
TRANSVERSE PROCESS ANGLE (TPA)						
Lee et al. 2021 [18]	Scolioscan	164	Manual	Thoracic and Lumbar	Thoracic: 3.0° (0°-9.9°) Lumbar: 2.8° (0°-11.9°)	Thoracic: 0.893 Lumbar: 0.884
Ungi et al. 2020 [32]	MicrUs EXT-1H	8	Automatic (U-Net segmentation)	Thoracic and Lumbar	2.2°	-
Brink et al. 2018 [58]	Scolioscan	33	Manual	Thoracic and Lumbar	4.7°± 3.6°	Thoracic: 0.992 Lumbar: 0.985

measurement, namely Cobb Angle [59], [60]; Spinous Process Angle (SPA) [17], [54], [58], [61], [62], [63], [64]; and Transverse Process Angle (TPA) [18], [32], [58]. Due to the absence of vertebral bodies in ultrasound images, the angle between vertebra endplates of a scoliotic curve i.e. Cobb angle, cannot be sufficiently measured [32] and the average correlation with radiographic Cobb Angle was 0.58 (blinded) and 0.85 (Aid Of previous Radiographs or AOR) [59]. Using posterior anatomical landmarks, alternate ways to measure scoliosis in ultrasound are being worked on. Two such measures are SPA and TPA. SPA is the angle formed between the lines drawn through the most tilted part of the spinous column profile of coronal ultrasound images while the TPA is measured from the lateral bony features of the spine. Both SPA and TPA have been proven comparable to traditional radiographic Cobb angle [32], [58]. However, the SPA measurement process has limitations in that the raters are disturbed by scattered and invalid points when manually tuning the inflection points of the spinous process curve, and thus the measurements of SPA were affected significantly in the lumbar area [54], [58]. In addition, SPA measurement uses different landmarks than the radiographic Cobb angle, introducing more variation sources for each measurement [18]. The variation between SPA and traditional Cobb angle is more pronounced for greater curvature of the spine [65]. The transverse process is an alternative way to assess scoliosis and the measuring index is called the transverse process angle (TPA). TPA shows an excellent correlation with radiographic Cobb angle [58]. In [32], the transverse process and ribs were segmented on 2-D transverse ultrasound images and reconstructed through 3-D reconstruction on the coronal plane. Measured from a 3D reconstructed image, their research gave a mean average difference (MAD) of 2° between X-ray-based angle and ultrasound TP angle and this difference is

TABLE 7
SUMMARY OF SEGMENTATION ARCHITECTURES APPLIED TO ULTRASOUND IMAGES OF SPINE AND OTHER BODY PARTS

Author	Field of work	Challenges	Seg. Arch. used	Special Features	Summary of dataset used	Dice Score	AUC	Jaccard/ mIOU
				OTHER BODY PARTS				
Amiri et al. 2020 [30]	Breast Lesion	Complexity of lesion shape and location	Two stage U-Net	one U-Net for ROI detection, one for segmentation	Training - 1398 , Testing - 1892, Total - 3290 Cross Validation: Five Fold	80.5	-	-
Chen et al. 2021 [31]	Foetal spina bifida	Large number of noise spot	Oct-U-Net	Octave feature to reduce redundant information	3,300 pregnant women's foetus images	-	-	91.7
				SPINE				
Ungi et al. 2020 [32]	Sagittal spine image	Automatic segmentation with a CNN and volume reconstruction	Basic U-Net	Basic U-Net features	No of Images: Training - 1398 , Testing - 1892, Total - 3290 Cross Validation: One Cross	-	97	-
Huang et al. 2020 [33]	Thoracic and lumbar bony features	Noise	RSN-U-net	Total variance loss to improve the robustness against the speckle and regular occlusion noise	No of Images: 109; Training - 80 , Testing - 29	78.38	98	-
Lyu et al. 2021 [34]	Thoracic and lumbar bony features	Noise, Variability of ROIs	D-TV Net	Two branches to estimate semantic region and contour segmentation, ASPP module to concatenate different scales of features	No of Images: 109;Training - 80 , Testing - 29, Cross Validation: 3 fold	86.68	-	-
Banerjee et al. 2021 [67]	Thoracic and lumbar bony features	Noise, Variability of ROIs	LDS U-Net	Light Dense Block to increase computation efficiency, Multiscale Skip-pathway to enhance feature fusion & Selection Gates to identify target bony features.	No of Images: 109;Training - 79 , Testing - 30, Cross Validation: 3 fold	86.94	-	-
Proposed Method	Thoracic and lumbar bony features	Noise, Variability of ROIs	SIU-Net	Improvised inception block to handle large variability in spine images & re-designed Dense-skip connection for multi-scale feature fusion	No of Images: 109;Training - 79, Testing - 30, Cross Validation: 5 fold	**88.3**	**99**	**78.0**

within the clinically significant error range. Ultrasound curvature angle (UCA) is a type of TPA, which is measured on coronal curvature on the AIS spine and uses thoracic and lumbar bony features as regions of interest [18]. Hence, given the advantages of TPA/UCA over SPA and ultrasound Cobb angle, this research is focused on automation of the UCA measurement process.

For automation of these measurement processes, research has been carried out to efficiently segment the relevant region of interest (ROI) from an ultrasound image as a step before the actual angle measurement. Ultrasound images have high speckle and scan noise which makes demarcation of necessary information challenging. For automatic SPA measurement, the segmentation was done for the mid spine line (single ROI) and its segmentation was done using a basic U-net [64]. On the other hand, for automatic UCA measurement, the proper identification and segmentation of thoracic and lumbar bony features are of prime importance. Single ROI segmentation techniques such as basic U-Net will not handle the variability in locations, sizes, and shapes of multiple ROI (TBFs and LBFs) in a noisy ultrasound image. Table 7 summarizes the performance of various segmentation architectures applied to ultrasound images of a breast lesion, fetal spina bifida, and spine.

In [33], an improvised version of U-Net i.e. U-Net with robustness to speckle and regular occlusion noise (RSN-U-Net) was introduced to segment lateral bony features from noisy ultrasound spine images. A new technique, i.e. total variance loss, was employed to improve the robustness against the speckle and regular occlusion noise. The research showed that RSN-U-Net only marginally outperformed basic U-Net with a segmentation result of 0.78 (Dice score) against the score of 0.76 by basic U-Net; indicating that the variability of location, shape, and sizes of bony features was not specifically and fully addressed through the RSN-U-Net architecture.

Lyu et al. employed a dual-task ultrasound transverse vertebrae segmentation network (D-TVNet) to segment lateral bony features from noisy ultrasound spine images [34]. The Atrous Spatial Pyramid Pooling (ASPP) [66] module was adopted to extract effective features from the spine images. An ASPP module is composed of four parallel atrous convolution layers with different dilated rates. Feature fusion was done by merging all the features extracted by each atrous convolution layer from different receptive fields. The four parallel atrous convolution layers in ASPP can achieve the effect equivalent to applying multiple filters with their different receptive fields on a given input image. In that work, the number of components in ASPP was increased by one by fusing another extra feature stream into ASPP concatenation. Three limitations could be observed: a) ASPP with limited sampling ranges will not be able to comprehensively extract the features of the target entities with variable sizes, b) some entities are so far from the ranges enclosed by the convolution kernels of ASPP that the features they have cannot be sampled, and c) to reduce the computational burden, all the points, but for the operative sampling points, in the convolution kernels are filled with zeros. This may result in a situation wherein, for the final result, the convolution kernel samples the information surrounding a particular pixel and may disregard the delicate local features corresponding to the positions with zeroes [66]. The sizes of the entities from which the ASPP module collect information, only vary within a limited range which, in the real scenario, will not suffice as the variability in location, shape, and sizes of entities would hinder the sampling of adequate information to generate the complete and precise features required for automatic UCA measurement.

Banerjee et al. presented a novel hybridized lightweight convolutional neural network architecture, called Light-Convolution Dense Selection U-Net or LDS U-Net [67]. This architecture had two main features – (a) Attention gates that improved the segmentation clarity by tackling the 'noisy' information and (b) Multi-scale skip-pathways replaced the conventional skip-pathways so that the problem of large variabilities in shape, size, and locations of the TBFs and LBFs could be handled. However, attention gate [68] has a drawback, when used in basic U-Net architecture, it requires a significantly more number of parameters but produces only marginal improvement in segmentation output.

A novel hybridized CNN architecture, multi-scale feature fusion skip-inception U-Net or SIU-Net is introduced in this research for ultrasound spine segmentation. 109 2D ultrasound spine images and their expert suggested truth masks are used as the input image dataset. Basic U-Net was taken as the starting architecture [28]. Basic U-Net gave segmentation output with missed and conjoint bony features (Avg. Jaccard: 0.709 and avg. Dice Score: 0.817). After analyzing the output images, it was deduced that several bony features got missed out due to the lack of flexibility of U-Net to choose multiple sizes of filters. As the first modification to increase the segmentation performance, the option of selecting multiple sizes of filters was incorporated with the basic U-Net architecture and the network was called the Inception+U-Net (IU) model. Instead of unnecessarily adding more layers in the basic U-Net, an improvised version of the inception block was used to

replace the conventional convolution operation. As the inception model was equipped with convolutional layers of varying kernel sizes, it was anticipated that this concept would solve the variability issues. However, after evaluation of the segmentation output, it was seen that the IU model did not perform satisfactorily, especially in and around the noisy areas (avg. Jaccard: 0.711 and avg. Dice Score: 0.846).

Further analysis showed a high degree of inconsistency between the features passed from the encoder network and features transmitted through the decoder network. The combination of two inconsistent sets of features caused incongruity during the learning of the network and affected the segmentation result [35]. The conventional skip path of the IU model was then replaced by the Residual path [35] to resolve the two incompatible sets of features in encoder and decoder sides and the model was named as Inception+Res path (IRs) model. This model performed better (avg. Jaccard: 0.734 and avg. Dice Score: 0.855) than the IU model as some further processing was incorporated in skip connections to make the feature maps from both sides more consistent. With this result, it was inferred that the IRs model could perform better in scenarios with less noise and clear visibility. However, as the input dataset is noisy ultrasound spine images,

TABLE 8
PROCESS FLOW OF AUTOMATIC UCA MEASUREMENT

	Step	Objective	Research output
1	Selection of image with best lateral features	To automatically select the images with best lateral features from all depths of 3D volume projection imaging	Convolution RankNet [69](Past work)
2	Identification of TBFs and LBFs	To overcome the challenges of the noisy image and variability of locations, shapes, and sizes of bony features and adequately detect the bony features through ultrasound spine image segmentation	SIU-Net (Current Work)
3	Final Scoliosis assessment	To find key points from detected bony features, calculate the main thoracic and thoraco-lumbar angles and compare against existing techniques.	Future work

single-scale feature fusion was insufficient for this segmentation work. A dense network was needed to perform multi-scale feature fusion to extract deeper layer features and fuse with the features from shallower layers. Then, the residual skip paths were replaced with encoder side dense skip connection [38] and it generated promising segmentation output. The new model was named as Inception + encoder skip path (ISP) model. As ultrasound image noise is the biggest challenge for this research, the ISP model, in many cases, failed to provide a clear segmentation edge boundary (avg. Jaccard: 0.757 and avg. Dice Score: 0.876). Therefore, for more semantically rich feature extraction and fusion, the ISP model skip connection was replaced with a decoder side dense skip connection. This model, SIU-Net, finally solved the problem of variability of bony features and speckle noise with better identification and segmentation of bony features with proper bone edge boundary detection (avg. Jaccard: 0.78 and avg. Dice Score: 0.883).

TABLE 9
COMPARISON OF DETECTION RATES FOR VARIOUS SEGMENTATION ARCHITECTURES IN ULTRASOUND SCOLIOSIS MEASUREMENT

Author	Method	Avg Detection Rate (%)
Huang et al. 2020 [33]	RSN-U-net	69.34
Lyu et al. 2021 [34]	D-TV Net	75.29
Banerjee et al. 2021 [67]	LDS U-Net	76.59
Proposed method	SIU-Net	**81.20**

The proposed network shows a promising segmentation output compared to baseline network U-Net [28] and advanced networks, such as MultiResUNet [35] and Unet++ [38]. The performance of SIU-Net is evaluated quantitatively using three popular indices, i.e. Jaccard index, Dice coefficient, and histogram Euclidean distance. Each time, SIU-Net outperforms all the other popular segmentation techniques giving the maximum Jaccard index (0.78) and Dice coefficient (0.883) and minimum histogram Euclidean distance (0.011) in comparison with baseline network U-Net, MultiResUNet and Unet++. This is because an inception block, by its nature of construction, is more adept in extracting features from different locations, shapes, and sizes than conventional convolution layers. Also, closer examination of the individual segmentation result, for some special cases, points to the fact that the SIU-Net does a superior job of identifying and segmenting the TBF and LBF pairs than the other networks (Fig 8, 9, 10 & 11). Finally, though MultiResUNet and Unet++ provide promising segmentation output in a few individual cases, the result demonstrates that the proposed network is more consistent in giving the adequate segmentation output for the overall range of input images.

In the manual UCA process, the evaluators must locate the proper points for line placements on the ultrasound images.

Identifying the feature below and above the T12 level depends on the expertise and judgment of the evaluators [18]. The overall process flow of automatic UCA measurement is shown in Table 8. First, the best quality images are selected based on the quality of the image using convolution RankNet [69]. The second key element in the automatic UCA measurement is identifying and segmentation of the lateral bony features (thoracic and lumbar) as they play a pivotal role in angle calculation. This can be measured using an index called the bony feature detection rate and a comparison of detection rates between other contemporary segmentation techniques is presented in Table 9. It can be concluded that SIU-Net gives the best detection rate not only against general segmentation architectures such as U-Net, U-Net++, and MultiResUNet (Fig 13) but also outperforms other architectures dedicated specifically to ultrasound spine images (Table 9). The future work of this research is to calculate the UCA angle automatically from segmented images and validate it against existing techniques.

The main limitation of this research is the low sample size of the image dataset. Though the segmentation results are promising, the generalization of performance needs to be tested in larger and more diverse patient populations. Another limitation of this work is that segmentation could not be done properly for almost 20% of images (shown in Fig 13). In such cases, the proposed method fails to differentiate the bony features or completely misses some features. Also, as a general observation, the segmentation result of LBFs is lesser consistent than that of TBFs. For this, the authors postulate that rescanning will be required for the images with missing or conjoined bony features for proper feature identification. Going forward, the automatic segmentation can be combined with the scanning process so that the scanner can get real-time feedback to conduct a targeted rescan of the unclear areas, if necessary. Image quality could be further improved by developing a flexible or small probe with good penetration to overcome the limitations of ultrasound imaging. Using different ultrasound machines, imaging protocols, and different sonographers will help improve the richness of input information for future studies.

VI. Conclusion

In this paper, a novel hybridized CNN architecture, SIU-Net, is proposed to overcome the segmentation challenges in ultrasound images that are full of speckle noise and low contrast. The proposed network employs an improvised inception block to overcome the challenges of the variability of locations, shapes, and sizes of bony features in spine image segmentation. The decoder side dense skip pathways help to do multi-scale feature fusion to improve segmentation accuracy in noisy ultrasound spine images. The performance of the proposed network is evaluated in terms of both the pictorial quality and segmentation accuracy using ultrasound VPI images. The results determine that this network gives an improved performance in both cases, which makes the network usable as a previous step of automatic UCA measurement. A future study can be conducted with a larger set of patient data to generalize the performance of the SIU-Net. The performance of the proposed method can be enriched by improving the image quality by developing a flexible or small probe with good penetration, making SIU-Net a viable preceding step for automatic scoliosis curvature angle measurement.

Conflict of Interest

No benefits in any form have been or will be received from a commercial party related directly or indirectly to the subject of this manuscript.

Acknowledgment

The project is partially supported by Hong Kong Research Grant Council, Research Impact Fund (R5017-18).

References

[1] Kim H, Kim HS, Moon ES, Yoon C-S, Chung T-S, Song H-T, et al. Scoliosis imaging: what radiologists should know. Radiographics. 2010;30:1823-42.

[2] Liu D, Yang Y, Yu X, Yang J, Xuan X, Yang J, et al. Effects of specific exercise therapy on adolescent patients with idiopathic scoliosis: A prospective controlled cohort study. Spine. 2020;45:1039.
[3] Tsiligiannis T, Grivas T. Pulmonary function in children with idiopathic scoliosis. Scoliosis. 2012;7:1-6.
[4] Li S, Yang J, Zhu L, Li Y, Peng H, Lin Y, et al. Left ventricular mechanics assessed by 2-dimensional speckle tracking echocardiography in children and adolescents with idiopathic scoliosis. Clinical spine surgery. 2017;30:E381-E9.
[5] Cobb J. Outline for the study of scoliosis. Instr Course Lect AAOS. 1948;5:261-75.
[6] Simony A, Hansen EJ, Christensen SB, Carreon LY, Andersen MO. Incidence of cancer in adolescent idiopathic scoliosis patients treated 25 years previously. European Spine Journal. 2016;25:3366-70.
[7] Lai KK-L, Lee TT-Y, Lee MK-S, Hui JC-H, Zheng Y-P. Validation of scolioscan air-portable radiation-free three-dimensional ultrasound imaging assessment system for scoliosis. Sensors. 2021;21:2858.
[8] Hwang BY, Mampre D, Ahmed AK, Suk I, Anderson WS, Manbachi A, et al. Ultrasound in traumatic spinal cord injury: a wide-open field. Neurosurgery. 2021; 89(3):372-382.
[9] Tawfik NA, Ahmed AT, El-Shafei TE, Habba MR. Diagnostic value of spinal ultrasound compared to MRI for diagnosis of spinal anomalies in pediatrics. Egyptian Journal of Radiology and Nuclear Medicine. 2020;51:1-11.
[10] Zhang J, Cui X, Chen S, Dai Y, Huang Y, Zhang S. Ultrasound-guided nusinersen administration for spinal muscular atrophy patients with severe scoliosis: an observational study. Orphanet Journal of Rare Diseases. 2021;16:1-8.
[11] Kalagara H, Nair H, Kolli S, Thota G, Uppal V. Ultrasound imaging of the spine for central neuraxial blockade: a technical description and evidence update. Current Anesthesiology Reports. 2021:1-14.
[12] Gai S, Zhang B, Yang C, Yu L. Speckle noise reduction in medical ultrasound image using monogenic wavelet and Laplace mixture distribution. Digital Signal Processing. 2018;72:192-207.
[13] Jiang WW, Zhou GQ, Lai K-L, Hu SY, Gao QY, Wang XY, et al. A fast 3-D ultrasound projection imaging method for scoliosis assessment. Mathematical Biosciences and Engineering. 2019; 16(3):1067-1081.
[14] Cheung C-WJ, Zhou G-Q, Law S-Y, Mak T-M, Lai K-L, Zheng Y-P. Ultrasound volume projection imaging for assessment of scoliosis. IEEE transactions on medical imaging. 2015;34:1760-8.
[15] Zhou G-Q, Zheng Y-P. Assessment of scoliosis using 3-D ultrasound volume projection imaging with automatic spine curvature detection. 2015 IEEE International Ultrasonics Symposium (IUS): IEEE; 2015. p. 1-4.
[16] Brignol A, Gueziri H, Cheriet F, Collins D, Laporte C. Automatic extraction of vertebral landmarks from ultrasound images: A pilot study. Computers in Biology and Medicine. 2020:103838.
[17] Zhou G-Q, Jiang W-W, Lai K-L, Zheng Y-P. Automatic measurement of spine curvature on 3-D ultrasound volume projection image with phase features. IEEE transactions on medical imaging. 2017;36:1250-62.
[18] Lee TT-Y, Lai KK-L, Cheng JC-Y, Castelein RM, Lam T-P, Zheng Y-P. 3D ultrasound imaging provides reliable angle measurement with validity comparable to X-ray in patients with adolescent idiopathic scoliosis. Journal of Orthopaedic Translation. 2021;29:51-9.
[19] Pandey PU, Quader N, Guy P, Garbi R, Hodgson AJ. Ultrasound bone segmentation: a scoping review of techniques and validation practices. Ultrasound in Medicine & Biology. 2020;46:921-35.
[20] Zhang J, Li H, Lv L, Zhang Y. Computer-aided Cobb measurement based on automatic detection of vertebral slopes using deep neural network. International journal of biomedical imaging. 2017; 2017:9083916.
[21] Krizhenvshky A, Sutskever I, Hinton G. Imagenet classification with deep convolutional networks. Proceedings of the Conference Neural Information Processing Systems (NIPS). p. 1097-105.
[22] Thong WE, Labelle H, Shen J, Parent S, Kadoury S. Stacked auto-encoders for classification of 3d spine models in adolescent idiopathic scoliosis. Recent Advances in Computational Methods and Clinical Applications for Spine Imaging: Springer; 2015. p. 13-25.
[23] Liu S, Wang Y, Yang X, Lei B, Liu L, Li SX, et al. Deep learning in medical ultrasound analysis: a review. Engineering. 2019;5:261-75.
[24] Hong R, Cheng W-H, Yamasaki T, Wang M, Ngo C-W. Advances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part III: Springer; 2018.
[25] Zeiler MD, Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision: Springer; 2014. p. 818-33.
[26] Kokabu T, Kanai S, Kawakami N, Uno K, Kotani T, Suzuki T, et al. An algorithm for using deep learning convolutional neural networks with three dimensional depth sensor imaging in scoliosis detection. The Spine Journal. 2021;21:980-7.
[27] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Medical image analysis. 2017;42:60-88.
[28] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation. International Conference on Medical image computing and computer-assisted intervention: Springer; 2015. p. 234-41.
[29] Liu W, Li W, Gong W. Ensemble of fine-tuned convolutional neural networks for urine sediment microscopic image classification. IET Computer Vision. 2019;14:18-25.

[30] Amiri M, Brooks R, Behboodi B, Rivaz H. Two-stage ultrasound image segmentation using U-Net and test time augmentation. International journal of computer assisted radiology and surgery. 2020;15:981-8.
[31] Chen L, Tian Y, Deng Y. Neural network algorithm-based three-dimensional ultrasound evaluation in the diagnosis of fetal spina bifida. Scientific Programming. 2021;2021:3605739.
[32] Ungi T, Greer H, Sunderland KR, Wu V, Baum ZM, Schlenger C, et al. Automatic spine ultrasound segmentation for scoliosis visualization and measurement. IEEE Transactions on Biomedical Engineering. 2020;67:3234-41.
[33] Huang Z, Wang L-W, Leung FH, Banerjee S, Yang D, Lee T, et al. Bone feature segmentation in ultrasound spine image with robustness to speckle and regular occlusion noise. arXiv preprint arXiv:201003740. 2020.
[34] Lyu J, Bi X, Banerjee S, Huang Z, Leung FH, Lee TT-Y, et al. Dual-task ultrasound spine transverse vertebrae segmentation network with contour regularization. Computerized Medical Imaging and Graphics. 2021;89:101896.
[35] Ibtehaz N, Rahman MS. MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation. Neural Networks. 2020;121:74-87.
[36] Bi L, Kim J, Kumar A, Fulham M, Feng D. Stacked fully convolutional networks with multi-channel learning: application to medical image segmentation. The Visual Computer. 2017;33:1061-71.
[37] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: A nested u-net architecture for medical image segmentation. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: Springer; 2018. p. 3-11.
[38] Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. IEEE transactions on medical imaging. 2019;39:1856-67.
[39] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, et al. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.
[40] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. Proceedings of the IEEE conference on computer vision and pattern recognition 2016. p. 2818-26.
[41] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, Inception-ResNet and the impact of residual connections on learning. Proceedings of the AAAI Conference on Artificial Intelligence. February 2017; pp. 4278-84.
[42] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. Proceedings of the IEEE conference on computer vision and pattern recognition2017. p. 4700-8.
[43] Zhang Z, Wu C, Coleman S, Kerr D. DENSE-INception U-net for medical image segmentation. Computer Methods and Programs in Biomedicine. 2020;192:105395.
[44] Badrinarayanan V, Kendall A, Cipolla R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence. 2017;39:2481-95.
[45]Abadi M, Barham P, Chen J, Chen Z, Davis A, Dean J, et al. Tensorflow: A system for large-scale machine learning. 12th USENIX symposium on operating systems design and implementation (OSDI'16) 2016. pp. 265-83.
[46] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:14126980. 2014.
[47] Jaccard P. The distribution of the flora in the alpine zone. 1. New phytologist. 1912;11:37-50.
[48] Dice LR. Measures of the amount of ecologic association between species. Ecology. 1945;26:297-302.
[49] Abbasian Ardakani A, Bitarafan-Rajabi A, Mohammadzadeh A, Mohammadi A, Riazi R, Abolghasemi J, et al. A hybrid multilayer filtering approach for thyroid nodule segmentation on ultrasound images. Journal of Ultrasound in Medicine. 2019;38:629-40.
[50] Korez R, Putzier M, Vrtovec T. A deep learning tool for fully automated measurements of sagittal spinopelvic balance from X-ray images: performance evaluation. European Spine Journal. 2020;29:2295-305.
[51] Horng M-H, Kuok C-P, Fu M-J, Lin C-J, Sun Y-N. Cobb angle measurement of spine from X-Ray images using convolutional neural network. Computational and Mathematical Methods in Medicine. 2019; 2019:6357171.
[52] Li H, Luo H, Huan W, Shi Z, Yan C, Wang L, et al. Automatic lumbar spinal MRI image segmentation with a multi-scale attention network. Neural Computing and Applications. 2021:1-14.
[53] Khandelwal P, Collins DL, Siddiqi K. Spine and individual vertebrae segmentation in computed tomography images using geometric flows and shape priors. Frontiers in Computer Science. 2021:66.
[54] Zheng Y-P, Lee TT-Y, Lai KK-L, Yip BH-K, Zhou G-Q, Jiang W-W, et al. A reliability and validity study for Scolioscan: a radiation-free scoliosis assessment system uliasing 3D ultrasound imaging. Scoliosis and spinal disorders. 2016;11:13.
[55] Huang Q, Deng Q, Li L, Yang J, Li X. Scoliotic imaging with a novel double-sweep 2.5-dimensional extended field-of-view ultrasound. IEEE transactions on ultrasonics, ferroelectrics, and frequency control. 2019;66:1304-15.
[56] Huang Q, Zeng Z, Li X. 2.5-D extended field-of-view ultrasound. IEEE transactions on medical imaging. 2017;37:851-9.
[57] Chen H, Zheng R, Qian L, Liu F, Song S, Zeng H. Improvement of 3D ultrasound spine imaging technique using fast reconstruction algorithm. IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control. 2021;68:3104-13.
[58] Brink RC, Wijdicks SP, Tromp IN, Schlösser TP, Kruyt MC, Beek FJ, et al. A reliability and validity study for different coronal angles using ultrasound imaging in adolescent idiopathic scoliosis. The Spine Journal. 2018;18:979-85.

[59] Zheng R, Young M, Hill D, Le LH, Hedden D, Moreau M, et al. Improvement on the accuracy and reliability of ultrasound coronal curvature measurement on adolescent idiopathic scoliosis with the aid of previous radiographs. Spine. 2016;41:404-11.
[60] Young M, Hill DL, Zheng R, Lou E. Reliability and accuracy of ultrasound measurements with and without the aid of previous radiographs in adolescent idiopathic scoliosis (AIS). European Spine Journal. 2015;24:1427-33.
[61] Zeng H, Zheng R, Le LH, Ta D. Measuring spinous process angle on ultrasound spine images using the GVF segmentation method. 2019 IEEE International Ultrasonics Symposium (IUS): IEEE; 2019. p. 1477-80.

[62] Zeng H-Y, Ge S-H, Gao Y-C, Zhou D-S, Zhou K, He X-M, et al. Automatic segmentation of vertebral features on ultrasound spine images using Stacked Hourglass Network. arXiv preprint arXiv:210503847. 2021.
[63] de Reuver S, Brink RC, Lee TT, Zheng Y-P, Beek FJ, Castelein RM. Cross-validation of ultrasound imaging in adolescent idiopathic scoliosis. European Spine Journal. 2021;30:628-33.
[64] Banerjee S, Ling SH, Lyu J, Su S, Zheng Y-P. Automatic segmentation of 3D ultrasound spine curvature using convolutional neural network. 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC): IEEE; 2020. p. 2039-42.
[65] Wu H-D, Liu W, Wong M-S. Reliability and validity of lateral curvature assessments using clinical ultrasound for the patients with scoliosis: a systematic review. European Spine Journal. 2020;29:717-25.
[66] Lian X, Pang Y, Han J, Pan J. Cascaded hierarchical atrous spatial pyramid pooling module for semantic segmentation. Pattern Recognition. 2021;110:107622.
[67] Banerjee S, Lyu J, Huang Z, Leung HFF, Lee TT-Y, Yang D, et al. Light-Convolution dense selection U-Net (LDS U-Net) for ultrasound lateral bony feature segmentation. Applied Sciences. 2021;11:10180.
[68] Oktay O, Schlemper J, Folgoc LL, Lee M, Heinrich M, Misawa K, et al. Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:180403999. 2018.
[69] Lyu J, Ling SH, Banerjee S, Zheng J, Lai K, Yang D, et al. Ultrasound volume projection image quality selection by ranking from convolutional RankNet. Computerized Medical Imaging and Graphics. 2021;89:101847.